sensors

MDPI

Article

On the Use of Deep Learning for Imaging-Based COVID-19 Detection Using Chest X-rays

Gabriel Iluebe Okolo [1,*], Stamos Katsigiannis [2], Turke Althobaiti [3] and Naeem Ramzan [1]

[1] School of Computing, Engineering and Physical Sciences, University of the West of Scotland, Paisley PA1 2BE, UK; Naeem.Ramzan@uws.ac.uk
[2] Department of Computer Science, Durham University, Durham DH1 3LE, UK; stamos.katsigiannis@durham.ac.uk
[3] Faculty of Science, Northern Border University, Arar 91431, Saudi Arabia; Turke.althobaiti@nbu.edu.sa
* Correspondence: Gabriel.Okolo@uws.ac.uk

Abstract: The global COVID-19 pandemic that started in 2019 and created major disruptions around the world demonstrated the imperative need for quick, inexpensive, accessible and reliable diagnostic methods that would allow the detection of infected individuals with minimal resources. Radiography, and more specifically, chest radiography, is a relatively inexpensive medical imaging modality that can potentially offer a solution for the diagnosis of COVID-19 cases. In this work, we examined eleven deep convolutional neural network architectures for the task of classifying chest X-ray images as belonging to healthy individuals, individuals with COVID-19 or individuals with viral pneumonia. All the examined networks are established architectures that have been proven to be efficient in image classification tasks, and we evaluated three different adjustments to modify the architectures for the task at hand by expanding them with additional layers. The proposed approaches were evaluated for all the examined architectures on a dataset with real chest X-ray images, reaching the highest classification accuracy of 98.04% and the highest F1-score of 98.22% for the best-performing setting.

Keywords: COVID-19; chest X-ray; deep learning; CNN; image classification

Citation: Okolo, G.I.; Katsigiannis, S.; Althobaiti, T.; Ramzan, N. On the Use of Deep Learning for Imaging-Based COVID-19 Detection Using Chest X-rays. *Sensors* **2021**, *21*, 5702. https://doi.org/10.3390/s21175702

Academic Editor: Paweł Pławiak

Received: 28 July 2021
Accepted: 20 August 2021
Published: 24 August 2021

1. Introduction

The 2019 novel corona virus (COVID-19) pandemic that was first reported in Wuhan, China, in December 2019, has become a public health issue around the world [1]. The infection that caused the COVID-19 pandemic was called a Severe Acute Respiratory Syndrome, also known as SARS-CoV-2 [2]. As of the second quarter of 2021, the COVID-19 pandemic keeps on affecting the well-being and health of the general public. A critical step in the battle against COVID-19 is a reliable and effective detection method for diagnosing infected patients, with the end-goal of prompt treatment and care. Corona viruses are a huge group of viruses that cause illness. Examples are Middle East Respiratory Syndrome (MERS-CoV), Severe Acute Respiratory Syndrome (SARS-CoV) and COVID-19. COVID-19's earliest symptoms include, fever, cough, fatigue or myalgia [3–5].

The main screening technique used for identifying COVID-19 cases is reverse transcription polymerase chain reaction testing (RT-PCR) [6,7], which can recognise SARS-CoV-2 RNA from respiratory samples gathered through various means, e.g., nasopharyngeal or oropharyngeal swabs. Initial findings have stated that RT-PCR testing shows relatively poor sensitivity [8]. Further findings showed that RT-PCR testing is highly specific and the probability of false positives is low. However, the amount of virus in a swab varies among patients, so at the initial test, it can provide a true negative result that turns out to be a false negative at a later stage, which is dangerous [9,10]. A screening method that can additionally be used for COVID-19 detection is radiography assessment, where chest radiography imaging such as computed tomography (CT) or chest X-ray (CXR) is conducted and analysed by radiologists to search for visual markers related to SARS-CoV-2

viral infection. Early investigations showed that patients present abnormalities in chest radiography images that pointed out features of those with COVID-19 [11,12], with some recommending that radiography assessment could be utilised as a primary tool for diagnosing COVID-19 [13]. However, the American College of Radiology (ACR) [14] and the World Health Organisation (WHO) [15] are sceptical and urge caution in using chest X-rays or chest CT scans as a primary diagnostic tool for COVID-19, with the WHO suggesting the use of chest imaging if RT-PCR is not available at all or in a timely manner.

Studies have shown that patients with COVID-19 exhibit some characteristics on their chest X-rays: It primarily affects the peripheral and lower areas of the lungs and presents nodular shadowing, ground glass opacity and accumulations of fluid and tissue in pulmonary alveoli, which is also called consolidation [16,17]. An important observation made during a study of COVID-19-related imaging diagnosis is that the initial symptoms are not visible at all or are slightly visible on chest X-rays within the first three days from symptom onset, but they are very obvious after 10 to 12 days [18]. A common complication of influenza-like illnesses is viral pneumonia, which has also been shown to be a complication of COVID-19 [19]. Medical imaging, specifically chest computed tomography (CT) and chest X-ray, is frequently utilised as an integrated assessment in the detection and management of pneumonia. However, given that viral pneumonia can be a complication of various illnesses, it is important to be able to assess whether a specific case is related to COVID-19.

The required medical and clinical resources for COVID-19 diagnosis at a global scale are a major challenge. Several countries are unable to carry out large numbers of COVID-19 tests [20], because of limited diagnosis tools. There is a need to identify a quick and reliable tool that can detect COVID-19 effectively with minimal effort. Numerous attempts have been conducted to devise an appropriate and quick approach to recognise infected patients at an early stage. After taking chest CT scans of 21 patients with COVID-19 in China, Guan et al. [21] found that CT scan analysis showed reciprocal pulmonary parenchymal abnormalities and pneumonic consolidation, as well as a fringe lung distribution. Thus, the analysis effectively extracted the main features of the virus.

All things considered, radiography assessment is quicker and has more prominent accessibility than RT-PCR testing, given the availability of chest radiology imaging systems in the healthcare sector. In addition, the turnaround time for X-ray examination is approximately 5.08 h [22]. Chest X-ray imaging is frequently used as a standard testing technique for respiratory complaints [23] and is easily accessible, making it a suitable COVID-19 detection method. Nevertheless, considering that COVID-19 symptoms are not visible in X-rays during the first days of the infection [18], chest X-ray imaging cannot fully replace RT-PCR, but can play an important role in patient screening to indicate potential COVID-19 cases, especially when RT-PCR is not easily accessible. However, probably the greatest bottleneck confronted is the requirement of expert radiologists to decipher the radiography images since the visual pointers can be unobtrusive. Consequently, computer diagnostic frameworks that can help radiologists quickly and precisely decipher radiography images to recognise COVID-19 cases are of critical importance for an accessible-to-all protect against the virus.

The critical need to develop solutions to help curb the challenges in the effort against COVID-19, motivated by the availability of CT and chest X-ray images of COVID-19 cases, led this study to carry out experiments on deep convolutional neural network (CNN) architectures that can effectively detect COVID-19 with the highest accuracy. To this end, we opted to select eleven well-established CNN architectures that have been shown to be efficient in various image classification tasks and conducted a comparative study to examine their performance on the task of classifying chest X-ray images as belonging to healthy individuals, individuals with COVID-19,or individuals with non-COVID-19-related viral pneumonia. Three different versions of the examined CNN architectures were evaluated: (1) a baseline version where the pretrained CNN models were used as is, changing only the classifier in their output to suit the task at hand; (2) a modified

version with two additional fully connected layers between the convolutional base and the classifier and dropout layers before each fully connected layer; and (3) a modified version with two larger fully connected layers between the convolutional base and the classifier and batch normalisation and leaky ReLU layers before each fully connected layer. All the examined approaches were trained and evaluated on a dataset with real chest X-ray images using a stratified five-fold cross-validation procedure, while the best-performing model was also evaluated on a completely unseen dataset. Supervised classification experiments demonstrated the efficiency of the proposed approach, reaching the highest classification accuracy of 98.04% and the highest F1-score of 98.22% for the best-performing setting.

The novelty of this work can be summarised as follows: (a) We provide a comparative study of the performance of multiple well-established CNN architectures that have been proven to work well on generic image classification tasks and examine them on the task of classifying chest X-ray images as belonging to healthy individuals, individuals with COVID-19 or individuals with non-COVID-19-related viral pneumonia. (b) We propose two different adjustments of these architectures and examine how classification performance is affected on the examined task. (c) We examine the performance of the models when using weights pretrained on ImageNet without any additional training and when end-to-end training is applied. (d) We show that although the pretrained CNN models have been proven to be very efficient on generic image classification tasks (e.g., ImageNet), performance suffers when fine-tuned for chest X-ray image classification, but minor extensions of the architectures, such as the ones proposed in this work, allow these CNN architectures to perform exceptionally well on the task, while also exploiting the available pretrained weights, thus reducing the amount of X-ray images needed for training the models. (e) Finally, similar available works in the literature commonly evaluate the performance of the proposed models by dividing their dataset into a training set and a test set. However, due to the limited availability of COVID-19-related images, such datasets are typically created by combining multiple datasets, which can lead to the trained neural networks learning features that are specific to the dataset better than the ones that are specific to the disease, thus leading to overfitting and reduced generalisation ability [24]. To ensure that the models proposed in this work do not suffer from this issue, in addition to our test set, we evaluated the performance of our models on a completely independent dataset, without any additional training or fine-tuning.

The rest of this paper is organised into five sections. Section 2 provides a brief literature review on the use of deep CNN architectures for medical image classification. Section 3 describes the proposed methodology, while the experimental results are presented and discussed in Section 4. Finally, conclusions are drawn in Section 5.

2. Related Work

The use of deep learning has proven to be very effective and reliable in revealing features, which are not evident, in images. Deep learning is currently widely used in the medical field for image classification and the detection of human diseases through computer-aided diagnosis [25–28]. Convolutional neural networks (CNNs) have demonstrated beneficial learning and useful feature extraction capabilities, and thus have been embraced by many researchers [29]. The use of pretrained networks on labelled medical images to train CNNs on disease classification has been proven to result in high performance, suggesting that in some cases, CNNs have achieved a level that is equivalent to or even better than certified human radiologists [30–32]. CNNs have been applied to recognise pulmonary nodules or masses from CT images [33], on chest X-ray images for the diagnosis of pneumonia [34], for cystoscopic image recognition extraction [35], for automated detection of polyps during colonoscopy [36], etc.

Transfer learning has also proven to be very efficient in deep learning applications, making it possible to use pretrained networks from different applications in order to save time and power in training a model to achieve high performance. This concept was utilised by Vikash et al. [37] in pneumonia detection utilising preprepared models trained

on the ImageNet [38] dataset. Xianghong et al. [39] modified the VGG16 model and used it for identification of lung regions and classification of various kinds of pneumonia. Ronneburger et al. [40] implemented a CNN with a small set of images, but applying a data augmentation technique to attain a better result. They applied the U-net to a cell segmentation task on two datasets, namely "PhC-U373" [41] and "DIC-HeLa" [42], and achieved an average IOU score of 92% and 77.5%, respectively. Ho et al. [43] reported an accurate identification of 14 thoracic diseases using feature extraction techniques and the pretrained DenseNet-121 [44] model. Lakhani et al. [31] also carried out an experiment on pulmonary TB detection using GoogLeNet [45] and AlexNet [29] by applying image augmentation techniques and attained an area under the curve (AUC) accuracy of 99%.

Wang et al. [46] carried out an experiment using chest X-ray data, labelled with eight diseases, and trained a deep CNN model by utilising weight parameters from VGGNet-16 [47], AlexNet [29], ResNet-50 [48] and GoogLeNet [45]. ResNet-50 showed better results than other models in the classification of seven diseases except for one, for which AlexNet performed better. Some of the AUC scores were as follows: "Cardiomegaly" (81.41%), "Pneumothorax" (78.91%), "Effusion" (73.62%), "Nodule" (71.64%), "Atelectasis" (70.69%).

Wang et al. [49] utilised deep learning methods on CT images to detect COVID-19 with a sensitivity, specificity and accuracy of 87%, 83% and 89.5%, respectively. Narin et al. [50] carried out an experiment on chest X-ray images, using Inception-ResNetV2 [51], InceptionV3 [52] and ResNet-50 [48] for the classification of COVID-19 and normal images. The ResNet50 model achieved the best classification accuracy of 98%, while for InceptionV3 97% and Inception-ResNetV2 87%. Wang et al. [53] presented a deep learning CNN architecture, called COVID-Net, for COVID-19 detection from chest X-rays, achieving a 92.4% accuracy.

Chowdhury et al. [54] carried out two COVID-19-related experiments with a chest X-ray dataset. The first utilised two classes, normal and COVID-19, while the second utilised three classes, namely COVID-19, viral pneumonia and normal. They experimented using transfer learning with and without image augmentation and tested and validated their approach using eight pretrained networks. Classification accuracy reached 99.41% for the two-class problem without image augmentation and 99.70% with image augmentation, while for the three-class problem, classification accuracy reached 97.74% without image augmentation and 97.94% with image augmentation.

The use of chest X-rays for COVID-19 detection has been the focus of multiple other recent studies. Shibly et al. [55] proposed the use of VGG16 [47] and the Faster R-CNN framework to detect COVID-19 from chest X-rays, achieving an accuracy of 97.36%, a sensitivity of 97.65% and a precision of 99.28%. Jain et al. [56] used the ResNet101 model, achieving a 98.93% accuracy, 98.93% sensitivity, 98.66% specificity, 96.39% precision and 98.15% F1-score. Nishio et al. [57] proposed a chest X-ray-based computer-aided diagnosis (CADx) system for classification into COVID-19 pneumonia, non-COVID-19 pneumonia and normal. They experimented using the VGG16 [47], MobileNet [58], DenseNet-121 [44], and EfficientNet [59] CNN models and reported that VGG16 performed best with an accuracy of 83.6%. Similarly, Apostolopoulos et al. [60] experimented with the VGG19 [47], MobileNetv2 [58], Inception [52], Xception [61] and InceptionResNetv2 [51] CNNs, reporting that MobileNetv2 performed best with an accuracy of 96.78%, a 98.66% sensitivity, and a 96.46% specificity. Sahlol et al. [62] attempted to reduce the computational complexity of CNN-based approaches by combining CNN-based features with the marine predators algorithm for swarm-based feature selection. Experiments on two different chest X-ray datasets demonstrated a maximum accuracy of 98.7%.

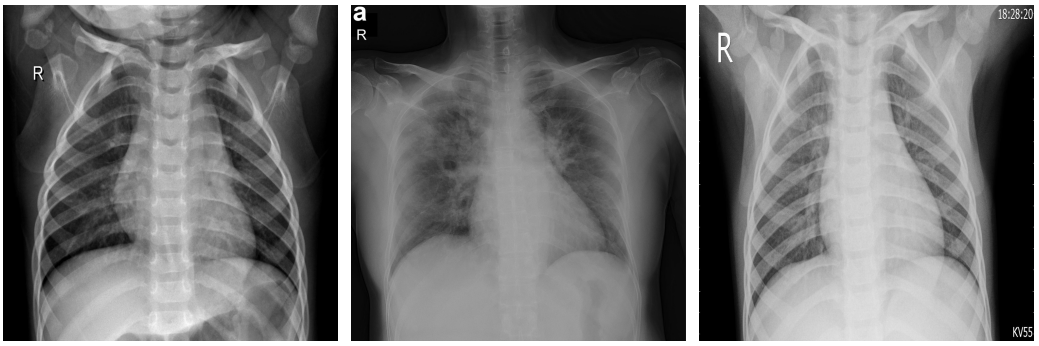
Apart from X-rays, other approaches have also been employed for COVID-19 detection. For example, considering that a cough is a vital symptom of COVID-19, Chuma et al. [63] carried out an experiment on a cough classification task using a K-band continuous-wave Doppler radar sensor and the AlexNet [29], VGG-19 [47] and GoogLeNet [45] CNN architectures, reporting that AlexNet performed best, with an accuracy of 88% when people were 1 m away from the sensor, 80% for 3 m and 86.5% for mixed 1 m and 3 m data.

3. Methodology

In this work, we examined various deep neural network architectures for the task of classifying chest X-ray images as belonging to healthy individuals, individuals with COVID-19 or individuals with viral pneumonia (non-COVID-19-related) and proposed three different adjustments to the architectures that led to increased performance for the task at hand. We opted to include viral pneumonia cases that were not related to COVID-19 as the third class in our experiments, since viral pneumonia is a complication of various diseases, but has been shown to also be a complication of COVID-19 [19]. The performance of the proposed approaches was evaluated on a publicly available chest X-ray image dataset [54], demonstrating their efficiency in improving COVID-19 detection regardless of the base network architecture used.

3.1. Dataset

The COVID-19 radiography database [54] was selected for this work. This database was compiled by a team of researchers from the University of Doha in Qatar and the University of Dhaka in Bangladesh, who collaborated with medical doctors from Pakistan and Malaysia to create a database of chest X-ray images for COVID-19-positive cases along with normal and viral pneumonia images. The database consists of 2905 chest X-ray images, including 219 COVID-19-positive images, 1341 normal images and 1345 viral pneumonia images. The images in the COVID-19 radiography database were collected from various sources, including the Italian Society of Medical and Interventional Radiology (SIRM) COVID-19 database [64], Cohen et al.'s COVID-19 image data collection [65], the ChestX-ray8 database [46], the Kermany et al. [66] pneumonia chest X-ray images dataset, as well as some online repositories [54], where physicians and researchers have uploaded COVID-19-related chest X-ray images. All the images are stored in Portable Network Graphics (PNG) file format (24 bit RGB), with a resolution of 1024 × 1024 pixels. Figure 1 depicts sample images from the database for COVID-19, normal and viral pneumonia chest X-ray images.



(**a**) Normal (**b**) COVID-19 (**c**) Viral pneumonia

Figure 1. Sample X-ray images from the used dataset.

3.2. Data Augmentation

One of the obstacles when attempting to apply deep learning techniques to solve a problem is the lack of sufficiently large amounts of data for training the deep learning models. Depending on the application and field, acquiring more data can be very arduous and costly, both in terms of time and resources. Data augmentation, i.e., increasing the amount of available data without gathering new data by applying various operations on the available data, has proven to be effective in image classification [67]. The technique has been used in the ImageNet classifier challenge by those that won the competition [29,48], and it is widely used by researchers to increase the training data, thereby avoiding overfitting [68].

In this work, we opted to use data augmentation techniques because of the limited number of images in the COVID-19 radiography database, especially for the COVID-19 class, which contained only 219 samples. To achieve this, the images in the training set at

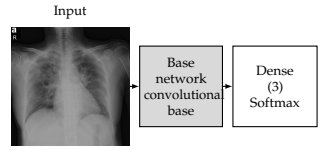
each training fold of the cross-validation procedure were used to create additional images by: randomly flipping them horizontally, randomly flipping them vertically, rotating at a random angle between −90 and 90 degrees, randomly shifting across the width by 10% of the total width, randomly shifting across the height by 10% of the total height, randomly zooming within a range of 0.9 to 1.1, randomly shearing in the counterclockwise direction by an angle of 0 to 0.1 rad, randomly shifting the brightness between 0.5 and 1.5, and finally, rescaling by a factor of 1/255. All pixels outside the boundaries of the input were filled using the nearest neighbour approach.

It must also be noted that all random values for the data augmentation operations were generated using a uniform probability distribution and that the Keras ImageDataGenerator class was used for the real-time creation of batches of augmented images during each training procedure. By using this data augmentation technique, the number of images in the training set was significantly increased, allowing the efficient use of deep learning techniques by training the machine learning models using a much larger amount of training images. Furthermore, it must be noted that the augmented images were only used for training the models and not for testing; thus, only original images from the dataset were used for testing the trained models.

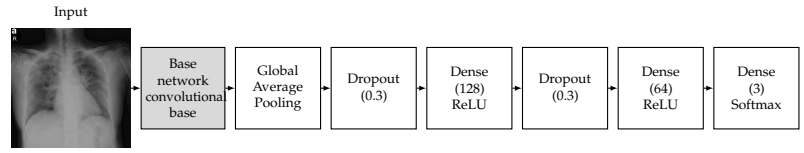
3.3. Classification Using Deep Neural Networks

Considering the limited amount of available images for the task at hand, we opted to base our approach on established architectures that have been proven to be efficient feature extractors for object detection applications and have been trained using sufficiently large image datasets. To this end, we selected eleven well-established CNN architectures that have achieved state-of-the-art results and were pretrained on the ImageNet [38] dataset, which consists of 1.4 million labelled images with 1000 classes. The selected deep learning models are very popular and widely used in computer vision tasks and have proven to excel in image classification problems. The following deep convolutional neural networks were examined for the task of classifying COVID-19, normal and viral pneumonia X-ray images (3-class problem): EfficientNetB4 [59], EfficientNetB7 [59], VGG16 [47], Xception [61], InceptionResNetV2 [51], InceptionV3 [52], MobileNetV2 [58], ResNet50V2 [48], DenseNet121, DenseNet169 and DenseNet201 [44]. The following three different approaches were used to adjust these architectures to the examined task, which were all trained using the Adam optimiser, a batch size of 16 and a learning rate of 0.0001. Furthermore, cross-entropy was used as the loss function, computed as $L_{CE} = -\sum_{c=1}^{M} y_{o,c} \log(p_{o,c})$, with $M = 3$ the number of classes, $y_{o,c}$ a binary indicator (0 or 1) if observation o belongs to class c and $p_{o,c}$ the predicted probability that observation o is of class c.

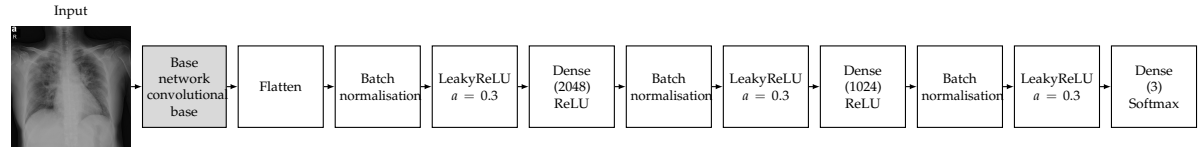
To train a robust CNN that will accurately classify images, hyperparameters must be tuned according to the examined problem. Our choice of hyperparameters was made after some preliminary experimentation and by following the conclusions of the study of Kandel et al. [69] on the effect of the batch size and learning rate when the Adam optimiser is used to train CNNs for medical image classification, which recommended that decreasing the batch size and lowering the learning rate will allow the network to learn better and generalise more accurately. It must also be noted that Keras and TensorFlow 2 were used for all the experiments; thus, all pretrained networks used in this work refer to the respective Keras implementations. Detailed diagrams of the proposed neural network architectures are depicted in Figure 2.



(a) Baseline approach



(b) Approach 1



(c) Approach 2

Figure 2. Proposed neural network architecture for the (**a**) baseline approach, (**b**) Approach 1 and (**c**) Approach 2.

3.3.1. Baseline Approach

The pretrained models were used as a feature extractor. CNNs are made up of two parts, which are the convolutional base and the classifier. The convolutional base contains the convolutional and pooling layers, which extract features from images. The classifier part is composed of a fully connected layer (softmax) with the goal of classifying images based on detected features. The concept of the baseline approach is to make use of only the convolutional base (feature extractor) without the classifier and feed its output directly into a softmax-activated layer [70] with 3 neurons (Figure 2a), corresponding to the number of targeted classes (COVID-19, normal, viral pneumonia). The convolutional base layer was set to freeze, in order to take advantage of the features learned by the models trained on the ImageNet dataset; therefore, the weights of the pretrained network were not updated during training. Then, the new classifier was trained to determine one of the three available classes given the set of extracted features [71]. It must be noted that contrary to the other examined architectures, VGG16 contains two fully connected layers before the final classifier. We opted to keep these layers in the VGG16 baseline model in order to be consistent with changing only the final softmax-activated layer for all the examined architectures.

3.3.2. Approach 1

The deep learning classification approach used is called round-off fine-tuning of the entire model. As shown in Figure 2b, for Approach 1, we added a new classifier with a new mini network of two small fully connected layers that fit our purpose. The first fully connected layer had 128 neurons, while the second had 64 neurons, followed by a softmax classifier with 3 neurons corresponding to our 3 output classes. We also added a global average pooling layer after the last convolutional block of the base network and a dropout layer with a rate of 0.3 before each of the two intermediate dense layers, which has been proven to help reduce the risk of overfitting [72]. Then, the pretrained weights on ImageNet were used as the initialisation of the base network in order to adapt the pretrained features to the new data.

3.3.3. Approach 2

The second approach is one of the most commonly used fine-tuning methods for image classification. We added a new classifier with two fully connected layers. The first fully connected layer had 2048 neurons, while the second had 1024 neurons, followed

by a softmax classifier with 3 neurons corresponding to our 3 output classes. As shown in Figure 2c, we also added a flatten layer after the output of the convolutional base of the base network and a batch normalisation layer before each of the three dense layers, which has also been proven to help reduce the risk of overfitting and accelerate the learning process [73]. We also opted to add a leaky ReLU layer after each batch normalisation layer, as this has be proven to improve the performance of a network [74]. Then, similar to Approach 1, the pretrained weights on ImageNet were used as the initialisation of the base network in order to adapt the pretrained features to the new data.

3.4. Hyperparameter Settings and Added Layers

Given the countless choices in the numbers, types and parameters of layers, we opted to examine the performance of architectures that were as simple as possible, adding only 3 dense layers, and compared them with a simpler approach (Approach 1) and a more complex one (Approach 2). The hyperparameters and added layers of the proposed approaches were selected by conducting some preliminary experiments on a smaller dataset, as follows: For all approaches, we opted to use the Adam optimiser, a batch size of 16 and a learning rate of 0.0001, as these settings performed best in a study carried out by Kandel et al. [69] on the effect of batch size and the impact of learning rates on the performance of CNNs for image classification of medical images. In addition, three fully connected (dense) layers were used after the convolutional layers of the base network for both Approaches 1 and 2. In both cases, the second dense layer had half the neurons of the first one, while the third dense layer (output layer) had three neurons corresponding to the three output classes.

For the dropout layers in Approach 1, a dropout rate of 0.3 was used. Dropout was proposed by Hinton et al. [72] as a regulariser that randomly sets a portion of the activations to the fully connected layers to zero during training, leading to improved generalisation ability and largely preventing overfitting [75]. Apart from the dropout layers, global average pooling was used in Approach 1 to generate a feature map from the output of the convolutional layers of the base network. It is a structural regulariser that helps to avoid overfitting in this layer, first proposed by Lin et al. [70].

Batch normalisation was used in Approach 2 to normalise activations in intermediate layers of the architecture, as it has been shown to improve accuracy, reduce the risk of overfitting and speed up the training process of deep neural networks [76]. Furthermore, a flatten layer was used in order to convert the multidimensional output of the convolutional layers of the base network into a one-dimensional vector, which can be fed into the fully connected layer [77]. Finally, Approach 2 used leaky ReLU activation layers ($\alpha = 0.3$), as they have been shown to improve the performance of a network by Wang et al. [74], who compared the performance of three different activation functions.

3.5. Label Smoothing

Label smoothing is a regularisation technique that addresses both overfitting and overconfidence problems. It is a simple method that makes a model more robust and enables it to generalise well. When cross-entropy is used as a loss function, the training process aims to minimise $L_{CE} = -\sum_{c=1}^{M} y_{o,c} \log(p_{o,c})$, where $y_{o,c}$ is a binary indicator (0 or 1) showing whether observation o belongs to class c. In this case, $y_{o,c}$ is considered a hard target as it is either 0 or 1. When label smoothing is used, the targets $y_{o,c}$ are modified as $y_{o,c}^{LS} = y_{o,c}(1 - \alpha) + \frac{\alpha}{M}$, with M being the number of classes and α the label smoothing parameter. Szegedy et al. [52] proposed the label smoothing technique, which improved the performance of the Inception architecture on the ImageNet dataset, and several other state-of-the-art deep learning classification models have adopted this method since [78,79]. In this work, we adopted label smoothing to improve the performance of our models by minimising cross-entropy using soft targets instead of hard targets, with a smoothing parameter $\alpha = 0.1$, thus encouraging the model to be less confident and leading to better generalisation.

4. Results and Discussion

The performance of the eleven examined deep convolutional neural network models for the there-class problem (normal, COVID-19, viral pneumonia) using the baseline and the other two proposed approaches was evaluated by conducting supervised classification experiments. Approach 1 and Approach 2 were evaluated twice, once keeping the pretrained weights of the base networks frozen and only training the additional layers and once using the pretrained weights as the initialisation and training the networks end-to-end. A stratified five-fold cross-validation procedure was followed in order to provide a fair estimate of the classification performance and avoid overfitting. To this end, the available images in the COVID-19 radiography database were divided into five groups, respecting the class distribution, and at each fold of the cross-validation procedure, one group was used for testing and the rest for training the examined models. This process was repeated until all groups had been used for testing, and the overall classification performance was computed by averaging the performance across the five folds.

The computed performance metrics were the accuracy, F1-score, precision and recall. Furthermore, since the F1-score, precision and recall depend on which class is considered as positive, their reported scores in this work are the average scores among the three examined classes. In addition to these four metrics, the Jaccard index and the Dice coefficient were also computed from the aggregated test groups across the five folds of the five-fold cross-validation procedure. All experiments were conducted by employing the TensorFlow library and the Keras API, using the Python programming language on the Google Colab Pro platform (Nvidia Tesla T4 and P100 GPU, 24 GB RAM). It must also be noted that the chest X-ray images were resized to 300×300 pixels before being fed as input to the examined network models, since this size was close to the input size for which they were originally designed. Given that the examined networks expected different input sizes and to achieve a fair comparison, we opted to resize the images to a common size that was close to the one expected by the networks, which varied from 224×224 to 456×456 pixels.

4.1. Results for the Baseline Approach

The classification performance achieved using the baseline approach and the two other proposed approaches is reported in Tables 1–3, respectively, in terms of the classification accuracy, F1-score, precision, recall, Jaccard index and Dice coefficient metrics. Table 1 contains the results for the baseline approach, which performed the worst among the examined approaches. The results were quite stable across all the proposed models, with an average F1-score within the range of 76.39–91.94%. VGG16 performed the best across all metrics, achieving the highest average F1-score of 91.94%, while EfficientNetB7 achieved the lowest average F1-score of 76.39%. The slightly higher performance of the VGG16 architecture compared to the others can be attributed to the additional two fully connected layers before the final softmax-activated layer, as explained in Section 3.3.1. Indeed, when performing the same experiment for VGG16 without the two fully connected layers, the F1-score dropped to 87.13%.

4.2. Results for Approach 1

Results for Approach 1 are reported in Table 2. From Tables 1–3, it is evident that Approach 1 with end-to-end training provided the best performance among the examined approaches, achieving considerably high metric values for all the examined models. In the case of end-to-end training, the EfficientNetB4- and Xception-based models achieved the highest average F1-scores of 98.22% and 98.20%, respectively, while the VGG16 and InceptionV3-based models achieved the lowest average F1-scores of 95.39% and 95.76%, respectively. On the other hand, the rest of the models, namely EfficientNetB7, ResNet50V2, InceptionResNetV2, MobileNetV2, DenseNet201, DenseNet169 and DenseNet121 models, achieved an average F1-score within the range of 96.66–97.92%. In the case of using the pretrained weights for the base network, performance suffered considerably compared to end-to-end training, with ResNet50V2 achieving the best performance across all metrics,

with an F1-score of 90.81%. Comparing the results from Tables 1 and 2, it is evident that the use of Approach 1 led to significant improvement in classification performance.

4.3. Results for Approach 2

Table 3 contains the classification results for Approach 2, which consistently performed slightly worse than Approach 1, as can be seen from Tables 2 and 3. In the case of end-to-end training, similar to Approach 1, the performance for Approach 2 was quite stable across the examined models, with an average F1-score within the range of 94.83–97.27%. The InceptionV3 and Xception models achieved the highest average F1-scores of 97.27% and 97.26%, respectively, while VGG16 achieved the lowest F1-score of 94.83%. In the case of using the pretrained weights for the base network, the performance decreased marginally compared to end-to-end training, with DenseNet169 and DenseNet121 achieving the highest F1-scores of 96.12% and 96.06%, respectively, with VGG16 achieving the lowest F1-score of 89.22%.

Table 1. Classification performance (%) of the examined deep neural network architectures following the baseline approach.

Base Model	Accuracy	F1-Score	Precision	Recall	Jaccard	Dice
EfficientNetB4	87.50	86.32	93.57	81.48	75.93	86.32
EfficientNetB7	83.75	76.39	91.45	70.86	69.40	81.94
VGG16	**92.22**	**91.94**	**93.97**	**90.13**	**83.21**	**90.84**
Xception	86.02	85.26	92.63	80.18	72.59	84.12
InceptionResNetV2	86.51	86.53	91.34	82.79	73.06	84.43
InceptionV3	86.82	84.79	90.73	80.58	72.60	84.12
MobileNetV2	86.68	85.76	91.87	81.33	72.76	84.24
ResNet50V2	89.02	88.90	93.56	85.14	77.11	87.08
DenseNet121	84.27	80.20	93.77	73.65	70.97	83.02
DenseNet169	87.23	86.52	92.84	82.03	74.75	85.55
DenseNet201	86.37	87.13	93.04	82.83	72.75	84.23

Note: Results in bold denote the best performance for each metric.

Table 2. Classification performance (%) of the examined deep neural network architectures following Approach 1 when using the pretrained weights for the base network and when training each network end-to-end.

Base Model	End-to-End Training						Pre-Trained Base					
	Accuracy	F1-Score	Precision	Recall	Jaccard	Dice	Accuracy	F1-Score	Precision	Recall	Jaccard	Dice
EfficientNetB4	**98.04**	**98.22**	98.52	**97.95**	**96.52**	**98.23**	89.95	88.71	92.86	85.55	80.35	89.10
EfficientNetB7	96.87	96.66	97.10	96.27	93.24	96.50	88.67	85.70	93.11	81.46	78.35	87.86
VGG16	94.77	95.39	96.11	94.84	88.83	94.08	87.57	85.59	89.23	82.86	72.85	84.29
Xception	98.00	98.20	**98.58**	97.87	95.98	97.95	90.36	90.07	92.29	88.16	79.92	88.84
InceptionResNetV2	97.38	97.35	97.88	96.89	94.17	97.00	89.05	88.76	91.25	86.65	77.51	87.33
InceptionV3	96.18	95.76	96.73	95.07	90.63	95.09	88.98	88.30	91.13	85.90	76.03	86.39
MobileNetV2	96.90	97.00	97.52	96.53	93.09	96.42	89.40	88.44	92.11	85.57	78.66	88.06
ResNet50V2	97.45	97.92	98.20	97.67	94.94	97.41	**91.02**	**90.81**	**93.43**	**88.68**	**81.23**	**89.64**
DenseNet121	97.07	97.46	97.95	97.03	93.83	96.82	88.06	87.28	93.23	83.02	75.66	86.14
DenseNet169	97.49	97.68	97.76	97.63	94.82	97.34	89.71	88.72	92.60	85.64	79.10	88.33
DenseNet201	97.25	97.15	97.60	96.77	93.93	96.87	89.12	88.25	91.63	85.58	78.21	87.77

Note: Results in bold denote the best performance for each metric and approach. Underlined results denote the overall best performance for each metric.

Table 3. Classification performance (%) of the examined deep neural network architectures following Approach 2 when using the pretrained weights for the base network and when training each network end-to-end.

Base Model	End-to-End Training						Pre-Trained Base					
	Accuracy	F1-Score	Precision	Recall	Jaccard	Dice	Accuracy	F1-Score	Precision	Recall	Jaccard	Dice
EfficientNetB4	96.45	96.60	97.38	95.86	92.45	96.08	93.22	92.12	95.46	89.61	86.13	92.55
EfficientNetB7	95.39	95.61	96.29	94.98	89.83	94.64	90.12	89.22	94.21	85.52	80.34	89.10
VGG16	94.66	94.83	95.61	94.15	88.70	94.01	93.94	93.74	94.40	93.13	86.28	92.64
Xception	96.72	97.26	97.80	**96.80**	93.31	96.54	93.49	93.90	94.62	93.29	86.39	92.70
InceptionResNetV2	96.70	96.12	95.83	96.58	91.68	95.66	93.94	94.02	95.63	92.57	86.23	92.60
InceptionV3	97.07	**97.27**	**98.14**	96.49	**94.59**	**97.22**	**96.18**	95.76	96.73	95.07	87.96	93.60
MobileNetV2	95.31	95.68	97.16	94.36	90.10	94.79	94.80	95.24	96.02	94.60	88.83	94.08
ResNet50V2	96.08	96.21	97.02	95.52	91.82	95.74	94.25	94.35	94.65	94.20	87.74	93.47
DenseNet121	96.18	96.55	96.88	96.25	92.89	96.31	95.42	96.06	**96.80**	95.44	90.64	95.09
DenseNet169	**97.11**	96.96	97.29	96.67	93.80	96.80	95.77	**96.12**	96.66	**95.64**	**91.17**	**95.38**
DenseNet201	96.52	96.73	97.05	96.51	92.64	96.18	95.04	95.02	95.86	94.26	89.15	94.26

Note: Results in bold denote the best performance for each metric and approach. Underlined results denote the overall best performance for each metric.

4.4. Validation on an Unseen Dataset

To further evaluate the generalisation ability of the proposed approach, we examined the classification performance of our best-performing model, i.e., the EfficientNetB4-based Approach 1, on an unseen dataset without any additional training or fine-tuning. To this end, we used the COVID-19 Image Repository (Version 2.0) [80], which contains 243 chest X-ray images of COVID-19 cases from the Institute for Diagnostic and Interventional Radiology, Hannover Medical School, Hannover, Germany. Considering that the available COVID-19 X-ray datasets are commonly collections of images from various sources, we selected this dataset in order to ensure that no overlap existed between its images and the images used for training our models (Section 3.1). As shown in Table 4, the EfficientNetB4-based Approach 1 model was able to correctly classify 234 out of the 243 (96.30%) COVID-19-related images, misclassifying 7 images as normal (2.88%) and 2 images as viral pneumonia (0.82%). These results on the unseen dataset that was not used for training or fine-tuning our model further demonstrated its efficiency and generalisation ability.

Table 4. Confusion matrix for the EfficientNetB4-based Approach 1 (end-to-end) without additional training or fine-tuning on the unseen COVID-19 dataset.

		Predicted		
		COVID-19	**Normal**	**Viral Pneumonia**
Actual	COVID-19	234 (96.30%)	7 (2.88%)	2 (0.82%)
	Normal	0 (0%)	0 (0%)	0 (0%)
	Viral pneumonia	0 (0%)	0 (0%)	0 (0%)

4.5. Execution Time

Information regarding the size in terms of trainable parameters and the time and number of epochs taken to train the best-performing models for the baseline approach, Approaches 1 and 2 with end-to-end training and Approaches 1 and 2 using the frozen pretrained weights for the base network is provided in Table 5. The average execution times were measured using TensorFlow and the Keras API on the Google Colab Pro platform (Nvidia Tesla T4 and P100 GPU, 24 GB RAM), as well as the training parameters described in Section 3.3. From Table 5, it is evident that the fastest model to train per epoch was VGG16 for the baseline approach (108 s/epoch), followed by the EfficientNetB4 and

Xception for Approach 1 with end-to-end training (112 and 111 s/epoch, respectively), which also achieved the best overall classification performance. DenseNet169 for Approach 2 using the pretrained weights for the base network required the most time to train at 166 s/epoch. However, it must be noted that for the sake of consistency and fairness, the training parameters were the same for all the models tested. Consequently, fine-tuning the parameters for each specific model could potentially lead to better overall training times for some of the models, and as a result, the execution times and number of epochs required for training that are reported in Table 5 must be taken into consideration only under the specific configuration and hardware.

Table 5. Execution time and size of the best-performing models.

Base Model	Approach	Trainable Parameters	Epochs	Training Time (s/epoch)
VGG16	Baseline	186,667,011	28	108 s
EfficientNetB4	Approach 1 (End-to-end)	17,786,571	32	112 s
Xception	Approach 1 (End-to-end)	21,077,675	20	111 s
ResNet50V2	Approach 1 (Pretrained base)	270,723	13	119 s
InceptionV3	Approach 2 (End-to-end)	28,076,195	18	123 s
Xception	Approach 2 (End-to-end)	27,114,795	21	118 s
DenseNet169	Approach 2 (Pretrained base)	5,520,643	18	166 s

Note: Average execution times measured using TensorFlow and the Keras API on the Google Colab Pro platform (Nvidia Tesla T4 and P100 GPU, 24 GB RAM). All base models were initialised with weights pretrained on the ImageNet dataset.

4.6. Discussion

From Tables 1–3, as well as from the the precision–recall plot in Figure 3, it is evident that both Approach 1 and Approach 2 led to improved performance compared to the baseline approach for all the examined base models. Furthermore, Approach 1 consistently provided higher classification performance, in terms of the F1-score, among all the approaches examined, regardless of the base model used. Consequently, Approach 1 demonstrated its superiority to modify pretrained image classification deep CNN models to classify chest X-ray images into normal, COVID-19 and viral pneumonia. Regarding the optimal base CNN model, the results were not conclusive for selecting a single model, but showed two out of the eleven examined candidates as suitable. The EfficientNetB4- and Xception-based models provided similar results for Approach 1 in terms of accuracy (98.04% vs. 98.00% respectively), F1-score (98.22% vs. 98.20%), precision (98.52% vs. 98.58%) and recall (97.95% vs. 97.87%), with the EfficientNetB4-based model being marginally better in most cases, while also achieved a marginally higher Jaccard index (96.52% vs. 95.98%) and Dice coefficient (98.23% vs. 97.95%). As a result, both models can be considered as suitable for the examined task.

Figure 4 depicts the aggregated confusion matrices, i.e., the sum of the confusion matrices from each fold of the five-fold cross-validation procedure, for Approach 1 (end-to-end training). It must be noted that since the additional images created through data augmentation (Section 3.2) were only used for training, the testing sets contained only the original images; thus, the number of samples for each class in the confusion matrices is equal to the number of samples per class in the dataset (Section 3.1). From these confusion matrices, it is evident that the two best-performing models achieved almost similar performance in correctly classifying COVID-19 samples, with the EfficientNetB4-based model correctly classifying 215/219 COVID-19 samples and the Xception-based model 214/219. Despite other models also achieving similar performance for the COVID-19 samples, the EfficientNetB4- and Xception-based models for Approach 1 achieved the best balance across all three available classes.

The EfficientNetB4 model belongs to the EfficientNet family, which consists of eight models, ranging from B0 to B7, and has been shown to achieve both higher accuracy and better efficiency than previous ConvNets, with a reduced parameter size. This group of models was developed by Google AI researchers and was scaled down by balancing the depth, width and resolution, which has led to effective results, and it is also smaller and faster than existing deep learning models [59]. More specifically, EfficientNetB4 achieved state-of-the-art 83.0% top-1/96.3% top-5 accuracy on ImageNet and has a size of 75 MB with over 19 million parameters [59]. The Xception model has previously achieved 79.0% top-1/94.5% top-5 accuracy on ImageNet and has a size of 88 MB with over 22 million parameters [61].

Interestingly, when using the pretrained weights for the base networks, Approach 2 outperformed Approach 1 (maximum F1-score of 96.12% for DenseNet169 vs. 90.81% for ResNet50V2, respectively). Despite performing worse than Approach 1 with end-to-end training, it seems that the more complex architecture of Approach 2 led to a better use of the pretrained features compared to the simpler architecture of Approach 1, when no end-to-end training was applied.

To further demonstrate the performance of the proposed approach, we compared the results of the best model for each approach (EfficientNetB4-based Approach 1 and InceptionV3-based Approach 2, both with end-to-end training) to other works in the literature that used the same dataset for the same classification task (COVID-19 vs. normal vs. viral pneumonia), as shown in Table 6. Given the large number of chest X-ray datasets in the literature and the fact that many works combine multiple datasets to increase the number of available samples, we opted to include in our comparison only works that used the exact same dataset as this work, in order to provide a fair comparison. From Table 6, it is evident that the proposed approach achieved a higher F1-score (98.22%) compared to the other methods, except for the CNN+BiLSTM approach of Aslan et al. [81], which achieved a marginally higher F1-score (98.76%) by utilising the more computationally complex BiLSTM layers on top of the CNN layers.

In addition, we used the gradient-weighted class activation mapping (Grad-CAM) method to visualise class activation heat maps for the best-performing model (EfficientNetB4-based Approach 1), as shown in Figure 5 for four COVID-19-positive images. The Grad-CAM method uses the gradients of any target class in a classification network flowing into the final convolutional layer to produce a coarse localisation map that highlights the most important image regions for predicting the specific class. It is based on the CAM method, which finds the discriminative regions for a CNN prediction through the computation of class activation maps, which assign importance to every position (i, j) in the last convolutional layer by computing the linear combination of the activations, weighted by the corresponding output weights for the observed class. Grad-CAM extends the CAM method by incorporating gradient information in the computation of the class activation maps (heat maps). By using the heat maps from the Grad-CAM method, we can examine the regions within the input image on which the CNN model focuses to make the decision for each class. By examining the examples in Figure 5 for our best-performing model, it is evident that the trained CNN model focuses on the areas of the lungs, as expected, and thus, we can be confident that the model relies on features extracted from the image regions that contain the information regarding COVID-19, viral pneumonia or healthy lungs, and not on information related to the images, to artefacts in the images or to the source of the images.

Regarding the applicability of our work in clinical practice, research has shown that the severity of COVID-19-related findings on chest X-rays peaks after 10–12 d from the initial onset of symptoms, whereas they are not visible or are slightly visible during the first 3 d [18]. Consequently, the proposed models could assist with COVID-19 diagnosis after symptom onset. However, it must be noted that the American College of Radiology (ACR) [14] and the WHO [15] urge caution in using chest radiography as a primary diagnostic tool for COVID-19, with the WHO suggesting its use in cases in which RT-PCR is

not available at all or in a timely manner. In addition, despite the high classification performance of the proposed models on images from various sources, a larger study that would include numerous COVID-19-positive chest X-rays, acquired from multiple radiography devices at multiple stages of the disease, would be required to evaluate their suitability for real-world clinical practice. Nevertheless, the acquired results are very promising, demonstrating how deep learning image classification models could potentially provide crucial help on the diagnosis of COVID-19.

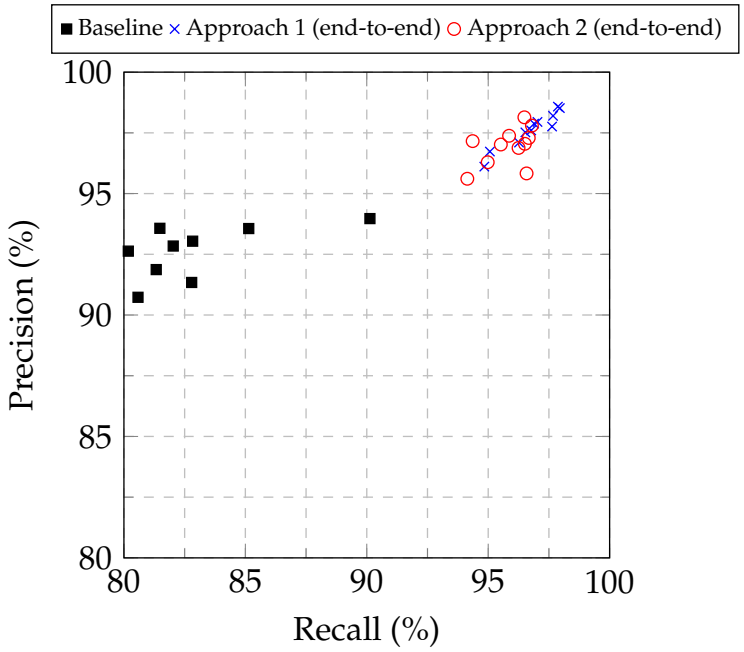


Figure 3. Precision vs. recall for the baseline approach and Approaches 1 and 2 for end-to-end training.

EfficientNetB4

Actual \ Predicted	COVID-19	Normal	Vir. Pneu.
COVID-19	215	3	1
Normal	1	1323	17
Vir. Pneu.	1	33	1311

EfficientNetB7

Actual \ Predicted	COVID-19	Normal	Vir. Pneu.
COVID-19	211	4	4
Normal	5	1313	23
Vir. Pneu.	8	44	1293

VGG16

Actual \ Predicted	COVID-19	Normal	Vir. Pneu.
COVID-19	208	4	7
Normal	13	1284	44
Vir. Pneu.	11	73	1261

Xception

Actual \ Predicted	COVID-19	Normal	Vir. Pneu.
COVID-19	214	2	3
Normal	2	1329	10
Vir. Pneu.	3	37	1305

InceptionResNetV2

Actual \ Predicted	COVID-19	Normal	Vir. Pneu.
COVID-19	210	4	5
Normal	4	1321	16
Vir. Pneu.	5	41	1299

InceptionV3

Actual \ Predicted	COVID-19	Normal	Vir. Pneu.
COVID-19	199	3	17
Normal	11	1288	42
Vir. Pneu.	3	35	1307

Mobile-NetV2

Actual \ Predicted	COVID-19	Normal	Vir. Pneu.
COVID-19	212	4	3
Normal	7	1315	19
Vir. Pneu.	9	45	1291

Res-Net50V2

Actual \ Predicted	COVID-19	Normal	Vir. Pneu.
COVID-19	215	3	1
Normal	3	1320	18
Vir. Pneu.	5	44	1296

Dense-Net121

Actual \ Predicted	COVID-19	Normal	Vir. Pneu.
COVID-19	214	3	2
Normal	4	1324	13
Vir. Pneu.	9	52	1284

Dense-Net169

Actual \ Predicted	COVID-19	Normal	Vir. Pneu.
COVID-19	216	3	0
Normal	7	1320	14
Vir. Pneu.	4	44	1297

Dense-Net201

Actual \ Predicted	COVID-19	Normal	Vir. Pneu.
COVID-19	210	6	3
Normal	1	1324	16
Vir. Pneu.	8	46	1291

Figure 4. Aggregated confusion matrices for Approach 1 (end-to-end training).

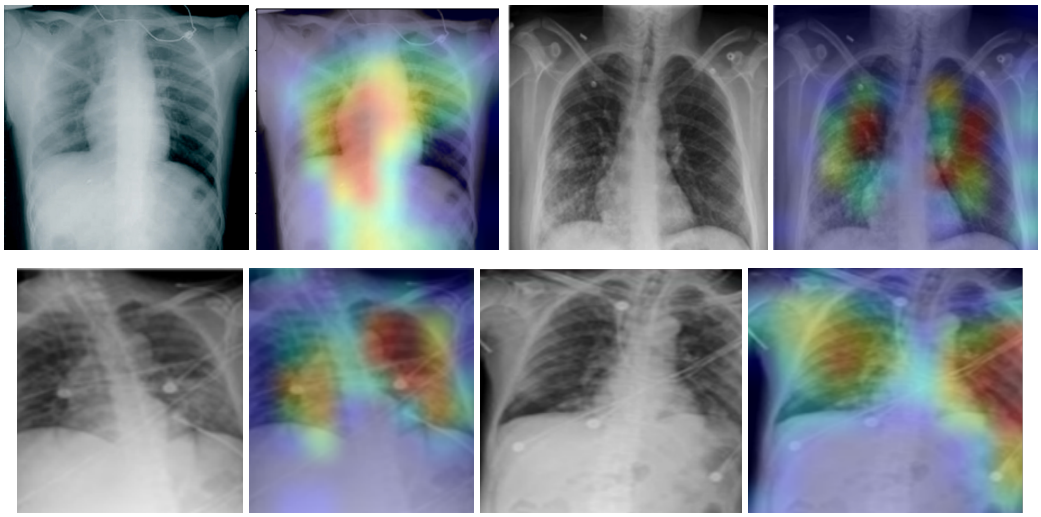


Figure 5. Grad-CAM visualisation for the COVID-19 class using the EfficientNetB4-based Approach 1 model for four chest X-ray images.

Table 6. Classification performance (%) of the best configurations versus other works using the same dataset for the same task.

Method	Accuracy	F1-Score	Precision	Recall
Aslan et al. [81] CNN	98.14	98.20	98.16	98.26
Aslan et al. [81] CNN+BiLSTM	98.70	98.76	98.77	98.76
Chowdhury et al. [54] (NIA)	97.74	96.61	96.61	96.61
Chowdhury et al. [54] (IA)	97.94	97.94	97.95	97.94
Maiti et al. [82] (GHE+TBH)	96.00	96.67	97.00	95.67
Öksüz et al. [83]	98.30	97.61	97.43	97.78
Progga et al. [84]	n/a	98.00	98.00	98.00
Sakib et al. [85]	96.00	97.67	98.00	97.67
This work-Approach 1	98.04	98.22	98.52	97.95
This work-Approach 2	97.07	97.27	98.14	96.49

NIA: no image augmentation, IA: image augmentation, GHE: global histogram equalisation, TBH: top bottom hat transform.

5. Conclusions

The global COVID-19 pandemic that started in 2019 has demonstrated the need for quick, inexpensive, accessible and reliable diagnostic methods for detecting infected individuals. In this work, we evaluated the performance of eleven deep convolutional neural network architectures for the task of classifying chest X-ray images as belonging to healthy individuals, individuals with COVID-19 or individuals with viral pneumonia. The eleven examined CNN models were selected due to their proven efficiency in image classification tasks and were modified in order to be adjusted for the task at hand. Supervised classification experiments using a five-fold cross-validation procedure were performed in order to evaluate the performance of three different modifications of the examined base CNN models on a dataset with real chest X-ray images that contained normal images, COVID-19-positive images and viral pneumonia images. The EfficientNetB4- and the Xception-based models, using Approach 1 and end-to-end training, provided the best classification performance, reaching an accuracy of 98.04% and 98.00%, respectively, and an average F1-score of 98.22% and 98.20%, respectively. Given the cost and accessibility of chest X-rays, the results achieved demonstrate the potential of the proposed approach for a relatively inexpensive and accessible diagnostic method for detecting COVID-19-positive individuals. Furthermore, the use of a dataset with images collected from various sources

indicates that the reported results are not constrained to a specific imaging device, but can be generalised, as also demonstrated by the very high accuracy (96.30%) achieved when classifying the images of an unseen dataset. Nevertheless, a larger study, including numerous COVID-19-positive chest X-rays, acquired from multiple radiography devices, would be required to evaluate the suitability of the proposed approach in real clinical practice.

To this end, future work will include a replication study using a much larger dataset of chest X-ray images, as well as a thorough study of the generalisation ability of the developed models by training and testing the models on diverse datasets from different sources and with X-ray images acquired by different X-ray machines. In addition, future work will also include a study on the explainability of the developed models, as well as on the use of the latest advances in deep-learning-based image classification, such as vision transformers.

Author Contributions: Conceptualisation, G.I.O., S.K., T.A. and N.R.; methodology, G.I.O., S.K., T.A. and N.R.; software, G.I.O.; validation, G.I.O., S.K., T.A. and N.R.; formal analysis, G.I.O., S.K., T.A. and N.R.; investigation, G.I.O., S.K., T.A. and N.R.; resources, N.R.; data curation, G.I.O.; writing—original draft preparation, G.I.O., S.K., T.A. and N.R.; writing—review and editing, S.K. and N.R.; visualisation, G.I.O. and S.K.; supervision, N.R. and S.K.; project administration, N.R. All authors read and agreed to the published version of the manuscript.

Funding: This research was partially supported by the SAFE_RH project at the University of the West of Scotland under Grant No. ERASMUS+ CBHE - 619483-EPP-1-2020-1-UK-EPPKA2-CBHE.

Institutional Review Board Statement: Ethical review and approval were waived for this study, due to all the data used being publicly available.

Informed Consent Statement: Not applicable.

Data Availability Statement: All datasets used in this work are publicly available from their respective sources.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

BiLSTM	Bidirectional long short-term memory
CNN	Convolutional neural network
COVID-19	Coronavirus Disease 2019
CT	Computed tomography
CXR	Chest X-ray
GHE	Global histogram equalisation
GPU	Graphics processing unit
Grad-CAM	Gradient-weighted class activation mapping
IA	Image augmentation
MERS-CoV	Middle East Respiratory Syndrome
NIA	No image augmentation
PNG	Portable network graphics
RAM	Random-access memory
ReLU	Rectified linear unit
RNA	Ribonucleic acid
RT-PCR	Reverse transcription polymerase chain reaction
SARS-CoV	Severe Acute Respiratory Syndrome
SARS-CoV-2	Severe Acute Respiratory Syndrome Coronavirus 2
TBH	Top bottom hat transform
WHO	World Health Organisation

References

1. Roosa, K.; Lee, Y.; Luo, R.; Kirpich, A.; Rothenberg, R.; Hyman, J.; Yan, P.; Chowell, G. Real-time forecasts of the COVID-19 epidemic in China from February 5th to February 24th, 2020. *Infect. Dis. Model.* **2020**, *5*, 256–263. [CrossRef] [PubMed]
2. Stoecklin, S.B.; Rolland, P.; Silue, Y.; Mailles, A.; Campese, C.; Simondon, A.; Mechain, M.; Meurice, L.; Nguyen, M.; Bassi, C.; et al. First cases of coronavirus disease 2019 (COVID-19) in France: Surveillance, investigations and control measures, January 2020. *Eurosurveillance* **2020**, *25*, 2000094.
3. Guan, W.J.; Ni, Z.Y.; Hu, Y.; Liang, W.H.; Ou, C.Q.; He, J.X.; Liu, L.; Shan, H.; Lei, C.L.; Hui, D.S.; et al. Clinical characteristics of coronavirus disease 2019 in China. *N. Engl. J. Med.* **2020**, *382*, 1708–1720. [CrossRef] [PubMed]
4. Wu, Z.; McGoogan, J.M. Characteristics of and important lessons from the coronavirus disease 2019 (COVID-19) outbreak in China: Summary of a report of 72314 cases from the Chinese Center for Disease Control and Prevention. *JAMA* **2020**, *323*, 1239–1242. [CrossRef]
5. Hope, M.D.; Raptis, C.A.; Henry, T.S. Chest Computed Tomography for Detection of Coronavirus Disease 2019 (COVID-19): Don't Rush the Science. *Ann. Intern. Med.* **2020**, *173*, 147–148. [CrossRef] [PubMed]
6. World Health Organization. *Clinical Management of Severe Acute Respiratory Infection when Novel Coronavirus (2019-nCoV) Infection Is Suspected: Interim Guidance, 28 January 2020*; Technical Documents; World Health Organization: Geneva, Switzerland, 2020.
7. Wang, W.; Xu, Y.; Gao, R.; Lu, R.; Han, K.; Wu, G.; Tan, W. Detection of SARS-CoV-2 in different types of clinical specimens. *JAMA* **2020**, *323*, 1843–1844. [CrossRef] [PubMed]
8. Fang, Y.; Zhang, H.; Xie, J.; Lin, M.; Ying, L.; Pang, P.; Ji, W. Sensitivity of Chest CT for COVID-19: Comparison to RT-PCR. *Radiology* **2020**, *296*, e115–e117. [CrossRef] [PubMed]
9. Wang, X.; Tan, L.; Wang, X.; Liu, W.; Lu, Y.; Cheng, L.; Sun, Z. Comparison of nasopharyngeal and oropharyngeal swabs for SARS-CoV-2 detection in 353 patients received tests with both specimens simultaneously. *Int. J. Infect. Dis.* **2020**, *94*, 107–109. [CrossRef]
10. Wikramaratna, P.S.; Paton, R.S.; Ghafari, M.; Lourenço, J. Estimating the false-negative test probability of SARS-CoV-2 by RT-PCR. *Eurosurveillance* **2020**, *25*, 2000568. [CrossRef] [PubMed]
11. Ng, M.Y.; Lee, E.Y.; Yang, J.; Yang, F.; Li, X.; Wang, H.; Lui, M.M.S.; Lo, C.S.Y.; Leung, B.; Khong, P.L.; et al. Imaging profile of the COVID-19 infection: Radiologic findings and literature review. *Radiol. Cardiothorac. Imaging* **2020**, *2*, e200034. [CrossRef] [PubMed]
12. Huang, C.; Wang, Y.; Li, X.; Ren, L.; Zhao, J.; Hu, Y.; Zhang, L.; Fan, G.; Xu, J.; Gu, X.; et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet* **2020**, *395*, 497–506. [CrossRef]
13. Ai, T.; Yang, Z.; Hou, H.; Zhan, C.; Chen, C.; Lv, W.; Tao, Q.; Sun, Z.; Xia, L. Correlation of Chest CT and RT-PCR Testing for Coronavirus Disease 2019 (COVID-19) in China: A Report of 1014 Cases. *Radiology* **2020**, *296*, e32–e40. [CrossRef] [PubMed]
14. American College of Radiology. ACR Recommendations for the Use of Chest Radiography and Computed Tomography (CT) for Suspected COVID-19 Infection. 2020. Available online: https://www.acr.org/Advocacy-and-Economics/ACR-Position-Statements/Recommendations-for-Chest-Radiography-and-CT-for-Suspected-COVID19-Infection (accessed on 1 June 2021).
15. Akl, E.A.; Blazic, I.; Yaacoub, S.; Frija, G.; Chou, R.; Appiah, J.A.; Fatehi, M.; Flor, N.; Hitti, E.; Jafri, H.; et al. Use of Chest Imaging in the Diagnosis and Management of COVID-19: A WHO Rapid Advice Guide. *Radiology* **2020**, *298*, e63–e69. [CrossRef] [PubMed]
16. Arentz, M.; Yim, E.; Klaff, L.; Lokhandwala, S.; Riedo, F.X.; Chong, M.; Lee, M. Characteristics and outcomes of 21 critically ill patients with COVID-19 in Washington State. *JAMA* **2020**, *323*, 1612–1614. [CrossRef] [PubMed]
17. Choi, H.; Qi, X.; Yoon, S.H.; Park, S.J.; Lee, K.H.; Kim, J.Y.; Lee, Y.K.; Ko, H.; Kim, K.H.; Park, C.M.; et al. Extension of Coronavirus Disease 2019 on Chest CT and Implications for Chest Radiographic Interpretation. *Radiol. Cardiothorac. Imaging* **2020**, *2*, e209004. [CrossRef]
18. Wong, H.Y.F.; Lam, H.Y.S.; Fong, A.H.T.; Leung, S.T.; Chin, T.W.Y.; Lo, C.S.Y.; Lui, M.M.S.; Lee, J.C.Y.; Chiu, K.W.H.; Chung, T.; et al. Frequency and distribution of chest radiographic findings in Patients Positive for COVID-19. *Radiology* **2020**, *296*, e72–e78. [CrossRef] [PubMed]
19. Heneghan, C.; Pluddeman, A.; Mahtani, K. Differentiating Viral from Bacterial Pneumonia. 2020. Available online: https://www.cebm.net/covid-19/differentiating-viral-from-bacterial-pneumonia/ (accessed on 1 June 2021).
20. Liu, H.; Liu, F.; Li, J.; Zhang, T.; Wang, D.; Lan, W. Clinical and CT imaging features of the COVID-19 pneumonia: Focus on pregnant women and children. *J. Infect.* **2020**, *80*, e7–e13. [CrossRef]
21. Guan, C.S.; Lv, Z.B.; Yan, S.; Du, Y.N.; Chen, H.; Wei, L.G.; Xie, R.M.; Chen, B.D. Imaging features of coronavirus disease 2019 (COVID-19): Evaluation on thin-section CT. *Acad. Radiol.* **2020**, *27*, 609–613. [CrossRef] [PubMed]
22. Albrecht, L.; Busse, R.; Tepe, H.; Poschmann, R.; Teichgräber, U.; Hamm, B.; de Bucourt, M. Turnaround time for reporting results of radiological examinations in intensive care unit patients: An internal quality control. *Radiologe* **2013**, *53*, 810–816. [CrossRef] [PubMed]
23. Nair, A.; Rodrigues, J.; Hare, S.; Edey, A.; Devaraj, A.; Jacob, J.; Johnstone, A.; McStay, R.; Denton, E.; Robinson, G. A British Society of Thoracic Imaging statement: Considerations in designing local imaging diagnostic algorithms for the COVID-19 pandemic. *Clin. Radiol.* **2020**, *75*, 329–334. [CrossRef] [PubMed]
24. Maguolo, G.; Nanni, L. A critic evaluation of methods for COVID-19 automatic detection from X-ray images. *Inf. Fusion* **2021**, *76*, 1–7. [CrossRef] [PubMed]

25. Liu, X.; Faes, L.; Kale, A.U.; Wagner, S.K.; Fu, D.J.; Bruynseels, A.; Mahendiran, T.; Moraes, G.; Shamdas, M.; Kern, C.; et al. A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: A systematic review and meta-analysis. *Lancet Digit. Health* **2019**, *1*, e271–e297. [CrossRef]
26. Kam, T.E.; Zhang, H.; Jiao, Z.; Shen, D. Deep Learning of Static and Dynamic Brain Functional Networks for Early MCI Detection. *IEEE Trans. Med. Imag.* **2020**, *39*, 478–487. [CrossRef] [PubMed]
27. Wang, Y.; Wang, N.; Xu, M.; Yu, J.; Qin, C.; Luo, X.; Yang, X.; Wang, T.; Li, A.; Ni, D. Deeply-Supervised Networks with Threshold Loss for Cancer Detection in Automated Breast Ultrasound. *IEEE Trans. Med. Imag.* **2019**, *39*, 866–876. [CrossRef] [PubMed]
28. Gordaliza, P.M.; Vaquero, J.J.; Sharpe, S.; Gleeson, F.; Munoz-Barrutia, A. A Multi-Task Self-Normalizing 3D-CNN to Infer Tuberculosis Radiological Manifestations. *arXiv* **2019**, arXiv:1907.12331.
29. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet Classification with Deep Convolutional Neural Networks. *Commun. ACM* **2017**, *60*, 84–90. [CrossRef]
30. Rajpurkar, P.; Irvin, J.; Zhu, K.; Yang, B.; Mehta, H.; Duan, T.; Ding, D.; Bagul, A.; Langlotz, C.; Shpanskaya, K.; et al. CheXNet: Radiologist-level pneumonia detection on chest X-rays with deep learning. *arXiv* **2017**, arXiv:1711.05225.
31. Lakhani, P.; Sundaram, B. Deep learning at chest radiography: Automated classification of pulmonary tuberculosis by using convolutional neural networks. *Radiology* **2017**, *284*, 574–582. [CrossRef] [PubMed]
32. Rajpurkar, P.; Irvin, J.; Bagul, A.; Ding, D.; Duan, T.; Mehta, H.; Yang, B.; Zhu, K.; Laird, D.; Ball, R.L.; et al. MURA: Large dataset for abnormality detection in musculoskeletal radiographs. *arXiv* **2017**, arXiv:1712.06957.
33. Choe, J.; Lee, S.M.; Do, K.H.; Lee, G.; Lee, J.G.; Lee, S.M.; Seo, J.B. Deep Learning–based Image Conversion of CT Reconstruction Kernels Improves Radiomics Reproducibility for Pulmonary Nodules or Masses. *Radiology* **2019**, *292*, 365–373. [CrossRef] [PubMed]
34. Nahid, A.A.; Sikder, N.; Bairagi, A.K.; Razzaque, M.; Masud, M.; Z Kouzani, A.; Mahmud, M. A Novel Method to Identify Pneumonia through Analyzing Chest Radiographs Employing a Multichannel Convolutional Neural Network. *Sensors* **2020**, *20*, 3482. [CrossRef]
35. Negassi, M.; Suarez-Ibarrola, R.; Hein, S.; Miernik, A.; Reiterer, A. Application of artificial neural networks for automated analysis of cystoscopic images: A review of the current status and future prospects. *World J. Urol.* **2020**, *38*, 2349–2358. [CrossRef] [PubMed]
36. Wang, P.; Xiao, X.; Brown, J.R.G.; Berzin, T.M.; Tu, M.; Xiong, F.; Hu, X.; Liu, P.; Song, Y.; Zhang, D.; et al. Development and validation of a deep-learning algorithm for the detection of polyps during colonoscopy. *Nat. Biomed. Eng.* **2018**, *2*, 741–748. [CrossRef] [PubMed]
37. Chouhan, V.; Singh, S.K.; Khamparia, A.; Gupta, D.; Tiwari, P.; Moreira, C.; Damaševičius, R.; de Albuquerque, V.H.C. A Novel Transfer Learning Based Approach for Pneumonia Detection in Chest X-ray Images. *Appl. Sci.* **2020**, *10*, 559. [CrossRef]
38. Russakovsky, O.; Deng, J.; Su, H.; Krause, J.; Satheesh, S.; Ma, S.; Huang, Z.; Karpathy, A.; Khosla, A.; Bernstein, M.; et al. ImageNet Large Scale Visual Recognition Challenge. *Int. J. Comput. Vis.* **2015**, *115*, 211–252. [CrossRef]
39. Gu, X.; Pan, L.; Liang, H.; Yang, R. Classification of bacterial and viral childhood pneumonia using deep learning in chest radiography. In Proceedings of the 3rd International Conference on Multimedia and Image Processing (ICMIP), Guiyang, China, 16–18 March 2018; pp. 88–93.
40. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In Proceedings of the 18th International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), Munich, Germany, 5–9 October 2015; pp. 234–241.
41. Dosovitskiy, A.; Springenberg, J.T.; Riedmiller, M.; Brox, T. Discriminative Unsupervised Feature Learning with Convolutional Neural Networks. In Proceedings of the 27th International Conference on Neural Information Processing Systems (NIPS), Montreal, QC, Canada, 8–13 December 2014; Volume 1, pp. 766–774.
42. Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Rich feature hierarchies for accurate object detection and semantic segmentation. In Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Columbus, OH, USA, 23–28 June 2014; pp. 580–587.
43. Ho, T.K.K.; Gwak, J. Multiple feature integration for classification of thoracic disease in chest radiography. *Appl. Sci.* **2019**, *9*, 4130. [CrossRef]
44. Huang, G.; Liu, Z.; Van Der Maaten, L.; Weinberger, K.Q. Densely connected convolutional networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 4700–4708.
45. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going deeper with convolutions. In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015.
46. Wang, X.; Peng, Y.; Lu, L.; Lu, Z.; Bagheri, M.; Summers, R.M. ChestX-ray8: Hospital-scale chest X-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 2097–2106.
47. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. In Proceedings of the 3rd International Conference on Learning Representations (ICLR), San Diego, CA, USA, 7–9 May 2015.
48. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.

49. Wang, S.; Kang, B.; Ma, J.; Zeng, X.; Xiao, M.; Guo, J.; Cai, M.; Yang, J.; Li, Y.; Meng, X.; et al. A deep learning algorithm using CT images to screen for Corona Virus Disease (COVID-19). *Eur. Radiol.* **2021**, *31*, 6096–6104. [CrossRef]
50. Narin, A.; Kaya, C.; Pamuk, Z. Automatic detection of coronavirus disease (COVID-19) using X-ray images and deep convolutional neural networks. *arXiv* **2020**, arXiv:2003.10849.
51. Szegedy, C.; Ioffe, S.; Vanhoucke, V.; Alemi, A.A. Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning. In Proceedings of the 31st AAAI Conference on Artificial Intelligence, San Francisco, CA, USA, 4–9 February 2017; pp. 4278–4284.
52. Szegedy, C.; Vanhoucke, V.; Ioffe, S.; Shlens, J.; Wojna, Z. Rethinking the Inception Architecture for Computer Vision. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 2818–2826. [CrossRef]
53. Wang, L.; Lin, Z.Q.; Wong, A. COVID-Net: A Tailored Deep Convolutional Neural Network Design for Detection of COVID-19 Cases from Chest X-Ray Images. *Sci. Rep.* **2020**, *10*, 19549. [CrossRef]
54. Chowdhury, M.E.H.; Rahman, T.; Khandakar, A.; Mazhar, R.; Kadir, M.A.; Mahbub, Z.B.; Islam, K.R.; Khan, M.S.; Iqbal, A.; Emadi, N.A.; et al. Can AI Help in Screening Viral and COVID-19 Pneumonia? *IEEE Access* **2020**, *8*, 132665–132676. [CrossRef]
55. Shibly, K.H.; Dey, S.K.; Islam, M.T.U.; Rahman, M.M. COVID faster R–CNN: A novel framework to diagnose novel coronavirus disease (COVID-19) in X-ray images. *Inform. Med. Unlocked* **2020**, *20*, 100405. [CrossRef]
56. Jain, G.; Mittal, D.; Thakur, D.; Mittal, M.K. A deep learning approach to detect COVID-19 coronavirus with X-Ray images. *Biocybern. Biomed. Eng.* **2020**, *40*, 1391–1405. [CrossRef]
57. Nishio, M.; Noguchi, S.; Matsuo, H.; Murakami, T. Automatic classification between COVID-19 pneumonia, non-COVID-19 pneumonia, and the healthy on chest X-ray image: Combination of data augmentation methods in a small dataset. *Sci. Rep.* **2020**, *10*, 17532. [CrossRef]
58. Sandler, M.; Howard, A.; Zhu, M.; Zhmoginov, A.; Chen, L.C. Mobilenetv2: Inverted residuals and linear bottlenecks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–23 June 2018; pp. 4510–4520.
59. Tan, M.; Le, Q.V. EfficientNet: Rethinking model scaling for convolutional neural networks. In Proceedings of the 36th International Conference on Machine Learning (ICML), Long Beach, CA, USA, 9–15 June 2019.
60. Apostolopoulos, I.D.; Mpesiana, T.A. Covid-19: Automatic detection from X-ray images utilizing transfer learning with convolutional neural networks. *Phys. Eng. Sci. Med.* **2020**, *43*, 635–640. [CrossRef]
61. Chollet, F. Xception: Deep learning with depthwise separable convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 1251–1258.
62. Sahlol, A.T.; Yousri, D.; Ewees, A.A.; Al-Qaness, M.A.; Damasevicius, R.; Abd Elaziz, M. COVID-19 image classification using deep features and fractional-order marine predators algorithm. *Sci. Rep.* **2020**, *10*, 15364. [CrossRef]
63. Chuma, E.L.; Iano, Y. A Movement Detection System Using Continuous-Wave Doppler Radar Sensor and Convolutional Neural Network to Detect Cough and Other Gestures. *IEEE Sens. J.* **2020**, *21*, 2921–2928. [CrossRef]
64. Italian Society of Medical and Interventional Radiology. COVID-19 Database. 2020. Available online: https://www.sirm.org/en/category/articles/covid-19-database/ (accessed on 1 June 2021).
65. Cohen, J.P.; Morrison, P.; Dao, L.; Roth, K.; Duong, T.Q.; Ghassemi, M. COVID-19 Image Data Collection: Prospective Predictions Are the Future. *arXiv* **2020**, arXiv:2006.11988.
66. Kermany, D.S.; Goldbaum, M.; Cai, W.; Valentim, C.C.S.; Liang, H.; Baxter, S.L.; McKeown, A.; Yang, G.; Wu, X.; Yan, F.; et al. Identifying medical diagnoses and treatable diseases by image-based deep learning. *Cell* **2018**, *172*, 1122–1131. [CrossRef]
67. Perez, L.; Wang, J. The effectiveness of data augmentation in image classification using deep learning. *arXiv* **2017**, arXiv:1712.04621.
68. Wong, S.C.; Gatt, A.; Stamatescu, V.; McDonnell, M.D. Understanding data augmentation for classification: When to warp? In Proceedings of the 2016 International Conference on Digital Image Computing: Techniques and Applications (DICTA), Gold Coast, QLD, Australia, 30 November–2 December 2016.
69. Kandel, I.; Castelli, M. The effect of batch size on the generalizability of the convolutional neural networks on a histopathology dataset. *ICT Express* **2020**, *6*, 312–315. [CrossRef]
70. Lin, M.; Chen, Q.; Yan, S. Network in network. In Proceedings of the International Conference on Learning Representations (ICLR), Banff, AB, Canada, 14–16 April 2014.
71. Chollet, F. *Deep Learning with Python*; Manning Publications: New York, NY, USA, 2018.
72. Srivastava, N.; Hinton, G.; Krizhevsky, A.; Sutskever, I.; Salakhutdinov, R. Dropout: A simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* **2014**, *15*, 1929–1958.
73. Ioffe, S.; Szegedy, C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In Proceedings of the 32nd International Conference on Machine Learning (ICML), Lille, France, 6–11 July 2015; pp. 448–456.
74. Wang, S.H.; Phillips, P.; Sui, Y.; Liu, B.; Yang, M.; Cheng, H. Classification of Alzheimer's disease based on eight-layer convolutional neural network with leaky rectified linear unit and max pooling. *J. Med. Syst.* **2018**, *42*, 85. [CrossRef]
75. Tetko, I.V.; Livingstone, D.J.; Luik, A.I. Neural network studies. 1. Comparison of overfitting and overtraining. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 826–833. [CrossRef]

76. Bjorck, J.; Gomes, C.; Selman, B.; Weinberger, K.Q. Understanding batch normalization. In Proceedings of the 32nd Conference on Neural Information Processing Systems (NIPS), Montreal, QC, Canada, 3–8 December 2018.
77. Ng, W.; Minasny, B.; Montazerolghaem, M.; Padarian, J.; Ferguson, R.; Bailey, S.; McBratney, A.B. Convolutional neural network for simultaneous prediction of several soil properties using visible/near-infrared, mid-infrared, and their combined spectra. *Geoderma* **2019**, *352*, 251–267. [CrossRef]
78. Zoph, B.; Vasudevan, V.; Shlens, J.; Le, Q.V. Learning transferable architectures for scalable image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–23 June 2018; pp. 8697–8710.
79. Real, E.; Aggarwal, A.; Huang, Y.; Le, Q.V. Regularized evolution for image classifier architecture search. In Proceedings of the AAAI Conference on Artificial Intelligence, Honolulu, HI, USA, 27 January–1 February 2019; Volume 33, pp. 4780–4789.
80. Winther, H.B.; Laser, H.; Gerbel, S.; Maschke, S.K.; Hinrichs, J.B.; Vogel-Claussen, J.; Wacker, F.K.; Höper, M.M.; Meyer, B.C. *Dataset: COVID-19 Image Repository*; Hannover Medical School: Hannover, Germany, 2020. [CrossRef]
81. Aslan, M.F.; Unlersen, M.F.; Sabanci, K.; Durdu, A. CNN-based transfer learning–BiLSTM network: A novel approach for COVID-19 infection detection. *Appl. Soft Comput.* **2021**, *98*, 106912. [CrossRef]
82. Maity, A.; Nair, T.R.; Chandra, A. Image Pre-processing techniques comparison: COVID-19 detection through Chest X-Rays via Deep Learning. *Int. J. Sci. Res. Sci. Technol.* **2020**, *7*, 113–123. [CrossRef]
83. Öksüz, C.; Urhan, O.; Güllü, M.K. Ensemble-CVDNet: A Deep Learning based End-to-End Classification Framework for COVID-19 Detection using Ensembles of Networks. *arXiv* **2020**, arXiv:2012.09132.
84. Progga, N.I.; Hossain, M.S.; Andersson, K. A Deep Transfer Learning Approach to Diagnose COVID-19 using X-ray Images. In Proceedings of the 2020 IEEE International Women in Engineering (WIE) Conference on Electrical and Computer Engineering (WIECON-ECE), Bhubaneswar, India, 26–27 December 2020.
85. Sakib, S.; Siddique, M.A.B.; Rahman Khan, M.M.; Yasmin, N.; Aziz, A.; Chowdhury, M.; Tasawar, I.K. Detection of COVID-19 Disease from Chest X-Ray Images: A Deep Transfer Learning Framework. *medRxiv* **2020**. [CrossRef]